

STUDY OF THE RARE DECAYS OF B_s^0 AND B^0 INTO MUON PAIRS WITH THE ATLAS DETECTOR

SANDRO PALESTINI

CERN, 1211 Geneva 23, Switzerland



on behalf of the ATLAS Collaboration

A study of the decays $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ has been performed using data corresponding to an integrated luminosity of 25 fb^{-1} of 7 TeV and 8 TeV proton–proton collisions collected with the ATLAS detector during the LHC Run 1. For B^0 , an upper limit on the branching fraction is set at $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% confidence level. For B_s^0 , the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$ is measured. The results are consistent with the Standard Model prediction with a p -value of 4.8%, corresponding to 2.0 standard deviations.

1 Introduction

The decays $B_{(s)}^0 \rightarrow \mu^+\mu^-$ are highly suppressed in the Standard Model (SM), as they are flavour-changing neutral-current (FCNC) processes, with additional helicity suppression. They are predicted accurately as $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10}$,¹ and deviations from the predicted values represent a sensitive test of new physics (*e.g.*, Ref. 2). The CMS and LHCb collaborations have reported the observation of $B_s^0 \rightarrow \mu^+\mu^-$ and evidence of $B^0 \rightarrow \mu^+\mu^-$, with combined values: $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}$.³ The results of a search performed with the ATLAS detector on the full sample of data collected during LHC Run 1 are reported here for the first time.^a

2 Analysis method, data samples, trigger and preliminary selection

In order to minimise systematic uncertainties, the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ branching fractions are measured relative to the normalisation decay $B^+ \rightarrow J/\psi (\rightarrow \mu^+\mu^-) K^+$, using the

^a More details can now be found in Ref. 4.

equation

$$BR(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = N_{d(s)} \times [\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)] \times \frac{f_u}{f_{d(s)}} \times \frac{1}{\mathcal{D}_{\text{norm}}},$$

with

$$\mathcal{D}_{\text{norm}} = \sum_{k=1,4} N_{J/\psi K^+}^k \alpha_k \left(\frac{\varepsilon_{\mu^+ \mu^-}}{\varepsilon_{J/\psi K^+}} \right)_k.$$

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are obtained from the number of decay candidates ($N_{d(s)}$), the known values of the B^+ branching fractions, the production ratio $B^+/B_{(s)}^0$ ($f_u/f_{d(s)}$), and a normalisation term $\mathcal{D}_{\text{norm}}$ that includes the yield of B^+ and the efficiency ratio between the search and the normalisation channel. This term is formed by four independent contributions, corresponding to the sample of data collected in 2011 and three exclusive trigger selections used in 2012. The coefficients α_k correct for the prescaling performed by the trigger selection. The notation used in the equation refers to both the stated and charge-conjugate process, without distinction between $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ and summing the $J/\psi K^+$ and $J/\psi K^-$ contributions. A blind analysis was performed in which data in the dimuon invariant mass region from 5166 to 5526 MeV were removed until the procedures for event selection and the details of signal yield extraction were completely defined.

A description of the ATLAS detector can be found in Ref. 5. The total integrated luminosity of good quality data used in this analysis is 4.9 fb^{-1} for the 2011 sample (collected at $\sqrt{s} = 7 \text{ TeV}$) and 20 fb^{-1} for 2012 ($\sqrt{s} = 8 \text{ TeV}$). Samples of simulated MC events are used for the determination of the efficiency ratios, for guiding the signal extraction fits, and for training and validation of the multivariate selections designed for the reduction of the background. In particular, a large sample of 1.4 billion inclusive dimuon events $pp \rightarrow b\bar{b}X \rightarrow \mu^+ \mu^- X'$ was used for the reduction of the combinatorial background. Samples of $B_{(s)}^0 \rightarrow h^+ h'^-$ were generated for the resonant background due to two-body decays with both hadrons misidentified as muons. Additional background samples were generated for partially reconstructed B -hadron decays (PRD) that contribute to the background to the $B_{(s)}^0$ and B^+ signals, including three-body semileptonic decays $h^- \mu^+ \nu$ with the hadron misidentified as a muon.

The online selection for the signal and the reference channels was based on dimuon trigger requirements. For the data collected in 2011 a threshold of 4 GeV was required on the transverse momentum (p_{T}) of both muons. The peak luminosity was higher in 2012 and, because of rate limitations, the same trigger was prescaled for part of the data taking. The loss of integrated luminosity was recovered to a large extent including two additional trigger selections, which require at least one of the muons in the barrel region (pseudorapidity $|\eta| < 1.05$) or above a higher threshold $p_{\text{T}} > 6 \text{ GeV}$.

Events compatible with $B_s^0 \rightarrow J/\psi \phi (\rightarrow \mu^+ \mu^- K^+ K^-)$ are also reconstructed, and used together with $B^+ \rightarrow J/\psi K^+$ to validate the simulation. The muon candidates were required to have tracks reconstructed in the inner detector and the muon spectrometer with a good matching between them. Kaons candidates are required to have $p_{\text{T}} > 1 \text{ GeV}$. For all tracks, the requirement of $|\eta| < 2.5$ is applied. The requirement of $p_{\text{T}} > 4 \text{ GeV}$ is placed on the B candidates.

The association between the decay vertex and a primary vertex is necessary for the final selection, discussed below. The matching is made by propagating the the B candidate to the point of closest approach to the collision axis, and choosing the primary vertex with the smallest longitudinal separation. Simulation shows that this method achieves a correct matching probability of better than 99%.

3 Background and multivariate selections

After the preliminary selection, the sample of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates is largely dominated by the combinatorial background due to uncorrelated decays of b and \bar{b} hadrons, denoted as the *continuum* background. A multivariate analysis, the *continuum-BDT*, is employed to enhance the signal relative to this background. This classifier is based on the 15 discriminating variables, which assess the characteristics of the decay vertex (*e.g.*: separation between decay and primary vertices, and collinearity between the separation between the two vertices and the momentum of the B candidate), the compatibility between the muon candidates and any reconstructed primary vertex, the compatibility of additional tracks with the B -meson decay vertex, and other general characteristics of the events, such as *isolation*, based on the comparison of the p_T of the B candidate with the sum of the p_T of the charged tracks reconstructed within a cone of size $\sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.7$ around the B direction. Figure 1-left shows the distribution of the BDT output variable for signal and background, separately for the continuum background and others sources. The classifier is trained on simulation, using the large sample of uncorrelated b decays containing muons, and the signal samples. The final selection requires a continuum-BDT output value larger than 0.24, corresponding to a signal relative efficiency of 54% and to a reduction of the continuum background by a factor of about 10^{-3} .

A second multivariate classifier, the *fake-BDT*, is used to enhance the signal against background containing hadrons misidentified as muons. This type of background represents a rather small fraction of the continuum and PRD components, but is relevant because of the resonant background of two-body hadron decays, in which both particles are misidentified as muons. Due to depth of the calorimeters and to the match between the tracks reconstructed in the inner detector and in the muon spectrometer, the hadronic punch-through is not relevant, and the *fake muons* are to a large extent due to decays in flight $K(\pi) \rightarrow \mu\nu$, when the muon takes almost all the meson momentum. In the p_T and η range of interest for this study, the misidentification probability for a kaon (pion) is about 0.4% (0.2%). This probability is further reduced applying the fake-BDT, which performs a multivariate classification on variables related to the quality of the reconstruction of the tracks in the inner detector and in the muon spectrometers, and to the match between the two. The misidentification probability is reduced by a factor 0.37, for a loss of genuine muons from B decays of 5%. Studies performed on data confirm the results of simulation within a factor 1.2 ± 0.2 , which is used as calibration and uncertainty. From the known values of the branching fractions $\mathcal{B}(B_{(s)}^0 \rightarrow h^+h'^-)$ and the efficiency obtained from simulation, the expected number of events from resonant background in the full sample of data is equal to 1.0 ± 0.4 events. A consistency check is performed inverting the selection applied with the fake-BDT. Doing so, the number of events containing real muons is largely reduced, while the number of resonant-background events is approximately three times larger than in the sample obtained with the nominal selection. A fit to the background-enhanced sample gives 0.5 ± 3.0 events of resonant background, in good agreement with the expectation.

4 Yield extraction for $B^+ \rightarrow J/\psi K^+$ and normalisation term

The B^+ yield for the normalisation channel is extracted with an unbinned extended maximum-likelihood fit to the $J/\psi K^+$ invariant mass distribution. The fit is performed separately on the four trigger/data categories, and performed on a prescaled sample of the data, corresponding to 2.7 fb^{-1} , collected and processed together with the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ candidates. The use of a reduced sample of $B^+ \rightarrow J/\psi K^+$ events does not affect the sensitivity of this study. Detailed functional forms, based on simulation, are used for the signal (including a non-Gaussian detector response, radiative tails, Cabibbo suppressed $B^+ \rightarrow J/\psi \pi^+$ component), and for the background (including a separate description of the main channels contributing to PRD). An example of the fit is shown in Figure 1-right. The total statistical and systematic uncertainty in the B^+ normalisation yield are in the range 0.9–4.0%, with the larger uncertainties

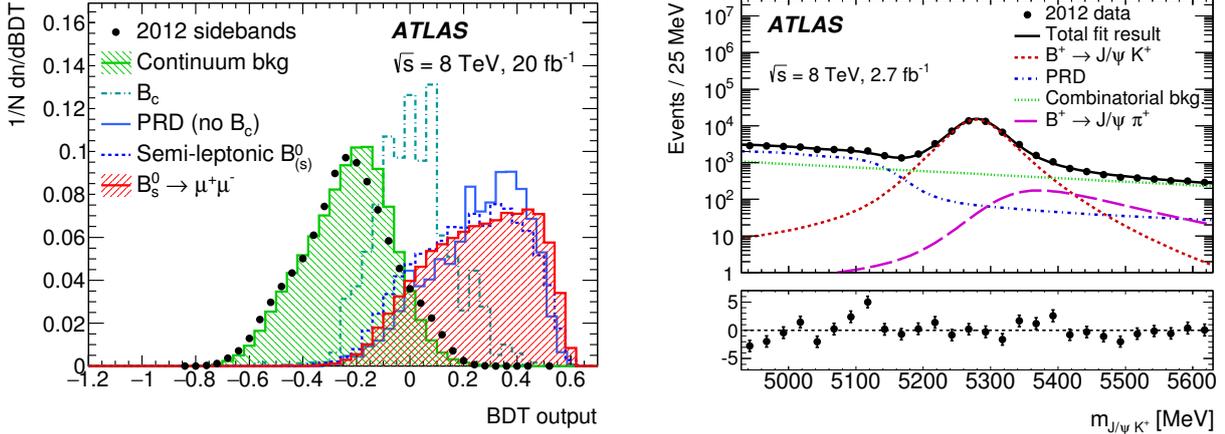


Figure 1 – *Left*: Continuum-BDT distribution for the $\mu^+\mu^-$ sideband data and MC samples of background (continuum, $B_c \rightarrow J/\psi\mu^+(\rightarrow\mu^+\mu^-\mu^+)$), partially reconstructed decays, $B_{(s)}^0 \rightarrow \pi^-(K^-)\mu^+\nu$ and signal. All distributions are normalised to unity. *Right*: $J/\psi K^+$ invariant mass distribution for B^+ candidates in the trigger category with the largest sample. The result of the fit is overlaid. The inset at the bottom shows the difference between the data point and the fit function, divided by the fit error.

in the categories that contribute less to the normalisation term $\mathcal{D}_{\text{norm}}$. The ratio of branching fractions $\mathcal{B}(B^+ \rightarrow J/\psi\pi^+)/\mathcal{B}(B^+ \rightarrow J/\psi K^+)$ is extracted from data, and found equal to $0.035 \pm 0.003 \pm 0.012$, in agreement with the world average.⁶

The efficiency ratio between the $B^+ \rightarrow J/\psi K^+$ and the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ channels is obtained from simulation, separately in each category of data, and combined with the B^+ yield to obtain the value of $\mathcal{D}_{\text{norm}}$. Simulation and data are compared using the background subtracted samples of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi\phi$, and the MC samples are tuned in order to reproduce more accurately the data. Corrections of 3–4% absolute size are due to (a) tuning to the p_T and η production spectrum of B mesons used in the generator, (b) use of trigger efficiencies extracted from studies performed on data, and (c) reweighting of the isolation variable, separately for B^+ and for $B_{(s)}^0 \rightarrow \mu^+\mu^-$, using $B_s^0 \rightarrow J/\psi\phi$ data for the latter. The total correction on $\mathcal{D}_{\text{norm}}$ is +3.2%. Finally, the decay time distribution of the $B_s^0 \rightarrow \mu^+\mu^-$ MC sample of is reweighted using the lifetime of the $B_{s,H}$ eigenstate of the $B_s^0\text{--}\bar{B}_s^0$ system, according to the SM based prediction for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays⁷ and the world average width difference $\Delta\Gamma_s$ ⁶ (–4% effect for B_s^0 , negligible for B^0). The value of $\mathcal{D}_{\text{norm}}$ is $(2.88 \pm 0.17) \times 10^6$ for B_s^0 and $(2.77 \pm 0.16) \times 10^6$ for B^0 .

5 Fit to event yields and branching fractions

The fit to the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ signal is an unbinned extended maximum-likelihood fit performed on the dimuon invariant mass distribution, with the events classified according to three intervals in the continuum-BDT output: 0.240–0.346, 0.346–0.446 and 0.446–1.

The model for describing signal and background is based on simulation and on data collected in the sidebands of the dimuon mass. The $B_{(s)}^0$ signal is described by the superposition of two Gaussian distributions, with a combined width of about 70 MeV. The efficiency of the BDT intervals is studied comparing data and simulation in the reference channels $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi\phi$. Relative corrections of +10% and –10% are applied to the first and third interval, together with systematic uncertainties of $\pm 14\%$, $\pm 6\%$ and $\pm 13\%$ respectively in the three intervals, due to MC inaccuracies in the modelling of the discriminating variables and of the continuum-BDT output.

The dependence of the continuum background on the dimuon mass is parameterised as a straight line, as observed in simulation and in good agreement with data in the sidebands. The

normalisation is determined independently in each BDT interval. The slope is extracted from the data, applying a loose Gaussian constraint of uniformity across the intervals.

The background due to partially reconstructed decays is described by an exponential dependence on the invariant mass. Both data and simulation show that the shape does not vary significantly across the BDT intervals. The normalisation in each interval and the common shape parameter are extracted from the fit to data. The mass distribution of the resonant background is determined from simulation, and is very similar to the B^0 signal. It is described with a Gaussian, with the normalisation equal to 1.0 ± 0.4 events, equally distributed among the three intervals of continuum-BDT.

Studies based on pseudo-experiments are used to assess the sensitivity to variations in the fit model. The effects considered include uncertainties in the reconstructed momentum scale and momentum resolution, in the mass dependence of the background components and in its variation across the continuum-BDT intervals. A direct modelling of semileptonic two-body decays $B_s^0(B^0) \rightarrow K(\pi)\mu\nu$ followed by $K(\pi)$ misidentification as a μ is included among the systematic studies. The total systematic uncertainty in the fit is equal to about ± 3 events for both B_s^0 and B^0 , with a correlation coefficient of -0.7 .

The branching fractions $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ are obtained by including in the fit: (a) the B^+ branching fraction to the final state (product of $(1.027 \pm 0.031) \times 10^{-3}$ and $(5.961 \pm 0.033) \times 10^{-2}$,⁶ (b) the production fractions, with $f_d/f_u = 1$ and $f_s/f_d = 0.240 \pm 0.20$,⁸ and (c) the normalisation term $\mathcal{D}_{\text{norm}}$, discussed above. The product of the three terms, *i.e.* the single-event sensitivity, is equal to $(8.9 \pm 1.0) \times 10^{-11}$ for $B_s^0 \rightarrow \mu^+\mu^-$ and $(2.21 \pm 0.15) \times 10^{-11}$ for $B^0 \rightarrow \mu^+\mu^-$. All systematic uncertainties are included in the likelihood as nuisance parameters with corresponding Gaussian constraints, and are treated with the profile-likelihood method.

6 Results

Including both the 2011 and 2012 data-taking periods, 1951 dimuon candidates are found in the full mass range of 4766–5966 MeV, including all three intervals of continuum-BDT. About 1680 events are determined as continuum background, 360 as PRD background, 16 ± 12 as $B_s^0 \rightarrow \mu^+\mu^-$ events, and -11 ± 9 as $B^0 \rightarrow \mu^+\mu^-$. The primary result of this analysis is obtained by applying the natural boundary of non-negative yields, for which the fit returns the central values 11 and 0 respectively for B_s^0 and B^0 .

Figure 2-left shows the dimuon mass distributions in the third BDT interval, together with the projection of the likelihood fit.

The result of the fit, applying the constraint of non-negative branching fractions, is:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (0.9_{-0.8}^{+1.1}) \times 10^{-9}, \quad \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10} \text{ (95\% CL)},$$

where the uncertainties include both the statistical and systematic contributions and are based on pseudo-experiments. The upper limit is obtained with the CL_s method. The observed significance of the $B_s^0 \rightarrow \mu^+\mu^-$ signal is equal to 1.4 standard deviations. The corresponding expected significance for the SM prediction is 3.1 standard deviations.

In addition to the central value and the 68.3% confidence range, an upper limit

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 3.0 \times 10^{-9} \text{ (95\%CL)}$$

is obtained. The limit is lower than the SM prediction, and compatible with the combination of the measurements by CMS and LHCb.³

The upper limit on $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ is above the SM prediction and also covers the central value of CMS and LHCb.

Pseudo-experiments are used to evaluate the compatibility of the observation with the SM prediction. A hypothesis test is performed for the simultaneous fit to both branching fractions. The result is $p = 0.048$, corresponding to 2.0 standard deviations.

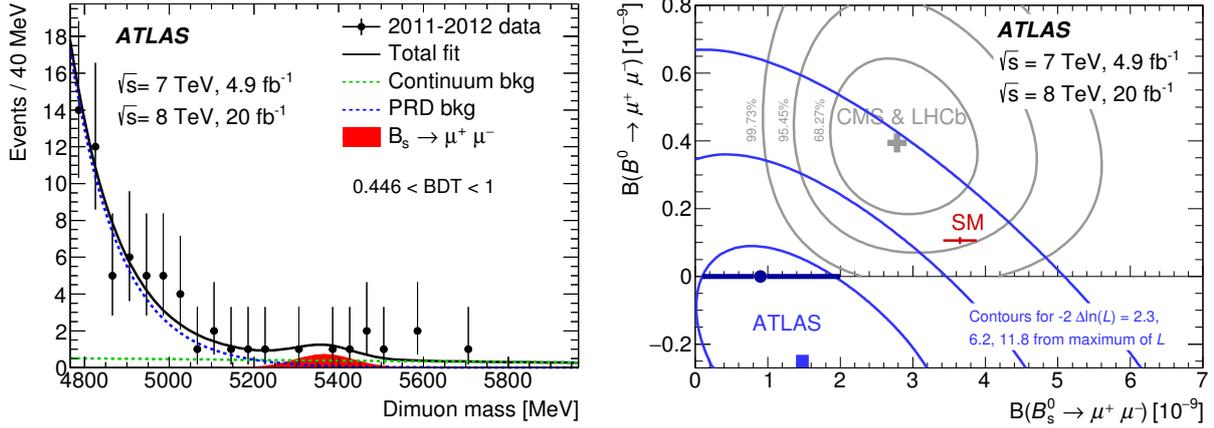


Figure 2 – *Left*: Dimuon invariant mass distributions in the third interval of continuum-BDT output, superimposed to the the maximum-likelihood fit. *Right*: Contours in the plane $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$, $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ for likelihood intervals relative to the absolute maximum. Also shown are the corresponding contours for CMS&LHCb, the SM prediction, and the maximum of the likelihood within the boundary of non-negative branching fractions, with the error bars covering the 68.3% confidence range for $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$.

Figure 2-right shows the contours in the plane of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ drawn for values of $-2\Delta\ln(L)$ equal to 2.3, 6.2 and 11.8, relative to the maximum of the likelihood, allowing negative values of the branching fractions. The maximum within the physical boundary is shown with error bars indicating the 68.3% interval for the value of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$. Also shown are the corresponding contours obtained in the combination of the results of the CMS and LHCb experiments, and the prediction based on the SM.

Acknowledgements

We thank CERN for the successful operation of the LHC, we acknowledge the support of the participating institutions and funding agencies, and of course the contributions of the collaborators in the ATLAS experiment. Special thanks to the small team of people directly committed to this challenging study.

References

1. C. Bobeth *et al*, *Phys. Rev. Lett.* **112**, 101801 (2012).
2. D.M. Straub in *Rencontres de Moriond EW 2012*, ed. E. Augé *et al* (ARISF, Paris, 2012).
3. CMS and LHCb Collaborations, *Nature* **522**, 68 (2015).
4. ATLAS Collaboration, arXiv:1604.04263 [hep-ex] (2016).
5. ATLAS Collaboration, *Journal of Instrumentation* **3**, S080003 (2008).
6. K.A. Olive *et al*, *Chin. Phys. C* **38**, 090001 (2014).
7. K. De Bruyn *et al*, *Phys. Rev. Lett.* **109**, 041801 (2012).
8. ATLAS Collaboration, *Phys. Rev. Lett.* **115**, 262001 (2016).